Informatics in Medicine Unlocked 19 (2020) 100326

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: http://www.elsevier.com/locate/imu

A multi-label learning model for psychotic diseases in Nigeria

S.O. Folorunso [a], S.G. Fashoto [b,*], J. Olaomi [c], O.Y. Fashoto [d]

[a] Department of Mathematical Sciences, Olabisi Onabanjo University, Ago-Iwoye, Ogun State, Nigeria
[b] Department of Computer Science, University of Eswatini, Kwaluseni, Eswatini, Swaziland
[c] Department of Statistics, University of South Africa, Florida Campus, South Africa
[d] Department of Psychology, Eswatini Medical Christian University, Mbabane, Eswatini

ARTICLE INFO

Keywords:
Mental illness
Psychotic disease
Multi-label classification
Label ranking
Binary relevance

ABSTRACT

The goal of Multi-Label Classification (MLC) is to allot an instance to a set of different labels. This task is usually addressed by either transforming the problem into several binary problems, adapting machine learning models to fit multi-label data or create an ensemble of models that can classify multi-label datasets. The communal relationship between Bipolar, Insomnia, Schizophrenia, Vascular Dementia (VD) and Attention-Deficit/Hyperactivity Disorder (ADHD) in the Psychotic Disorder Diseases (PDD) motivate the research for a diagnostic method that classifies and evaluates each psychotic disorder simultaneously.

This study experimentally evaluates 15 MLC methods using 10 evaluation measures over a new PDD dataset. The performance of these methods is measured with four ranking - based, three example-based and three label-based measures. Also, the efficiency of these methods is measured by their 90-10 Train-Test split with the 10 evaluation measures.

The results show that the Label Powerset (LC) and Pruned Sets (PS), MLC methods with Naïve Bayes (NB) and Naïve Bayes Tree (NBTree) consistently performed best in terms of the evaluation measures on the PDD dataset. Schizophrenia has the highest classification accuracy with Bipolar the lowest in the data split of 90-10. Logistic model tree (LMT) is the best algorithm for Insomnia and Bipolar while Naïve Bayes (NB) is the best for Schizophrenia, VD and MBD. Support vector machines (SVM) with ensemble learning and classification (ELC) and Ensemble of Pruned Set (EPS) are the best classifiers for Bipolar while SVM with regression and threshold (RT) is the least. The classifiers are statistically significantly different for Insomnia, VD and ADHD only.

1. Introduction

Psychotic Disorder Disease (PDD) is a well-researched area even with Machine Learning (ML) methods. But diagnosing these PDD simultaneously in patients are rare. The communal relationship between Bipolar, Insomnia, Schizophrenia, Vascular Dementia (VD) and Minimal Brain Dysfunction (MBD) motivate the research for a diagnostic method that classifies and evaluates each psychotic disorder simultaneously. PDD is a form of mental illness distinguished by loss of reality and entails observable symptoms [1]. Schizophrenia is a disorder linked to 'poor insight' of environ such as deterioration in social functions, disruptions in thoughts, emotion and language [2]. Insomnia is sleep-wake disorder while Bipolar is a subtype of mood disorder. Vascular dementia is a cerebrovascular disease that alters the normal functioning of blood vessels in the brain [3]. The Attention-Deficit/Hyperactivity Disorder formerly known as Minimal Brain Dysfunction (MBD), is described as chronic, pervasive developmental condition which entails problems with sustained attention, impulse control and activity regulation [4].

MLC is a well-heeded machine learning research area as labels in many real-world applications [5]. [6] reported the evaluation of multi-labeled medical data on depression and co-occurring obtained at the University of Benin Teaching Hospital (UBTH) and primary care centers in Nigeria with 1090 examples, 22 featured symptoms with 2-classes attributes. This dataset was evaluated with Bayesian Classifier Chains (BCC) [7], Probabilistic Classifier Chains (PCC) [8], Super class classifier (SCC) [9], Bagging [10], Ensemble of Classifier Chains (ECC) [11], Pruned Sets [12] and Classifier Chains (CC) [13] based on Hamming Loss, Hamming Score and Exact Match. BCC and PCC were found to perform well with 10-fold cross-validation [13]. proposed Multi-Label Problem Transformation Joint Classification (MLPTJC) model for a health and disease risk prediction dataset consisting of 110,300 instances, 22 features and 8 disease labels. They proposed a Joint

* Corresponding author.
E-mail address: sgfashoto@uniswa.sz (S.G. Fashoto).

https://doi.org/10.1016/j.imu.2020.100326
Received 18 December 2019; Received in revised form 24 March 2020; Accepted 29 March 2020
Available online 7 April 2020

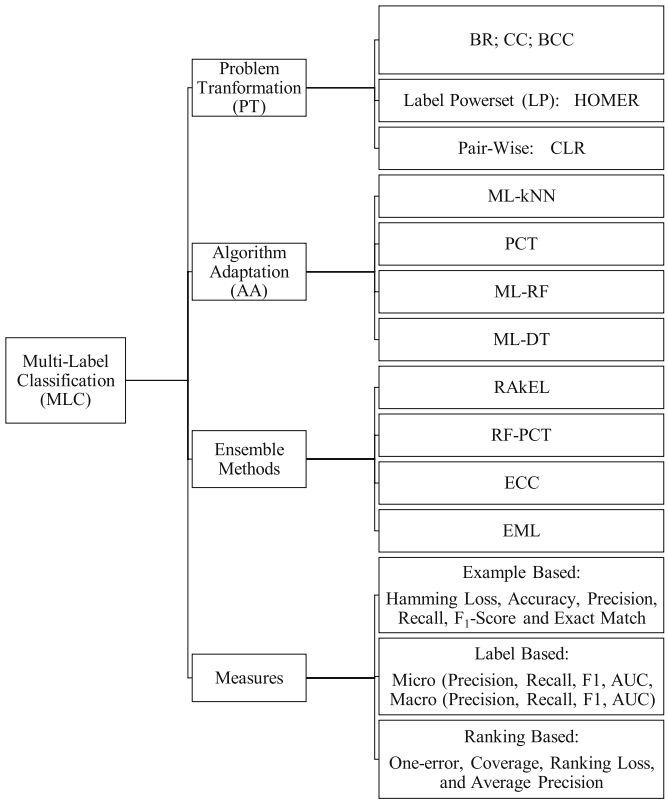

Fig. 1. Taxonomy of MLC [5].

Decomposition Subset Classifier method to reduce the infrequent label sets to deal with the imbalance learning problem. They applied the Support Vector Machine (SVM) and Random Forest (RF) models with 10-fold cross-validation with their performance measures based on Average Accuracy, Precision, Recall and F-measure. SVM gave a better performance. Likewise [14], reported findings on IncRNA multi label dataset with 89 composition features, 21 structural features of 7566 tissue-specific with 22 classes of tissues. BP-MLL, ML-kNN, BR and RAkEL methods were applied with SVM as base model. Based on the three-performance measure categorization, ML-kNN performed best [13]. proposed an Ensemble Label Power-set Pruned datasets Joint Decomposition (ELPPJD) model. The chronic disease risk dataset consisting of 110,300 instances, 62 symptoms and 6 disease labels (hypertension, diabetes, fatty liver, cholecystitis, heart disease, and obesity) were first transformed into a multiclass classification pruning the datasets and applying joint decomposition methods to deal with the imbalance learning problem. Size Balanced (SB) and Label Similarity (LS) decomposition methods were applied to the training dataset. ELPPJD was contrasted with RAkEL and HOMER and evaluated based on Average accuracy micro (Precision, Recall, and F1) and macro (Precision, Recall and F1). ELPPJD method with label similarity strategy has outstanding performance [14]. proposed a fusion of ML-kNN and Logistic Regression (LR) called IBLR-ML. This style captures the interdependencies between labels and combines model-based and similarity-based inference for multi-label classification. Based on Hamming Loss (HL), One Error (OE), Average Precision (AP), Rank Loss and Coverage, the proposed model outperforms BR and LP with k-NN, LR and Decision Tree (DT) as base models [29]. proposed a general framework to automatically transform a conventional instance representation into meta-level features, enables a broad range of learning-to-rank algorithms in information retrieval (IR) to be leveraged for category ranking in MLC, and invokes supervised learning for instance-based threshold optimization. The experiments were evaluated on multiple benchmark datasets and compared with Rank-SVM, ML-kNN, IBLR and Binary SVM. The proposed model significantly outperforms the other models at 5% significance level.

This study presents a new PDD dataset as a Multi-Label classification problem and experimentally assesses the data with 15 MLC methods using 10 evaluation measures. The performance of these methods is measured with four ranking-based, three example-based and three label-based measures. The efficiency of these methods is measured by their training and testing sets.

The study thus builds an MLC model, which uses MLC classifiers that provide consistent and clinician interpretable diagnostic results with respect to simultaneously determining the presence of the five PDD from symptoms accurately. It also comparatively evaluates four multi-label classification algorithms on the aforesaid dataset using a variety of evaluation metrics.

The remainder of the paper is organized as follows: Section 2 briefly exposes the subject of MLC and how it is related to PDD. Section 3 describes the methodology adopted in the study; Section 4 presents the results and its analysis while Section 5 concludes the study with a plan for future work.

2. MULTI-LABEL classification (MLC)

A Single - Label Classification (SLC) learns from a set of samples of a dataset associated with a single label λ from a set of disjoint labels L, $|L| > 1$. A binary classification problem occurs when $|L| = 2$, but becomes a multi-class problem when $|L| > 2$. But MLC occurs when samples of a dataset are associated with a set of labels $Y \subseteq L$ [5].

MLC models can be characterized into two groups namely Problem Transformation (PT) and Algorithm Adaptation (AA) methods as shown in Fig. 1.

PT methods are algorithm independent. Here, MLC task are transformed into one or more Single-label classification, regression or ranking tasks [5]. Some of the examples of models that fall under this group are: Binary Relevance (BR) [5], Classifier Chain (CC) [16], Bayesian Classifier Chain (BCC) [17], Label Power-set (LP), Pair-Wise: Calibrated Label Ranking (CLR) [7].

AA methods extends specific learning algorithms to treat multi-labeled data [5]. Some of the algorithms belonging to this category are: Multi-Label k Nearest Neighbour (ML-kNN) [18], ADABOOST.MH and ADABOOST.MR [19], Multi-Label C4.5 (ML-C4.5) [20], Predictive Clustering Trees (PCTs) [21] and Back Propagation- Multi Label Learning.

[22] extended the characterization of MLC to ensemble methods. This category comprises of methods that ensembles MLC models but uses either AA or PT as their base classifiers. Some examples of methods belonging to this category are Random k-label sets (RAkEL) [23], Ensembles of Classifier Chains (ECC) [19], Random Forests of Predictive Clustering Trees (RF-PCT) [24,25] and Random Forests of multi label C4.5 trees [21], Ensemble of Pruned Set (EPS) [12].

2.1. PDD as a Multi–Label classification problem

A Psychotic patient may also be suffering from other related diseases. The diagnosis of psychosis and related diseases (labels) from symptoms (attributes) is an MLC problem. Here, the many psychotic diseases, which are concurrently determined, explains the symptoms. These class/labels variables usually exhibit conditional dependence relations among themselves which must be modelled and learnt. Probability theory provides the framework for considering such possible multiple outcomes. Formally, the psychotic disease risk classification is framed into a multi-label classification problem. Given a set of r PDD records $T = \{m_1, m_2, \cdots, m_r\}$, with m_i $i = 1, 2, \cdots, r$and a set of n psychotic disease labels $L = l_1, l_2, \cdots, l_n$, *with* l_j, $j = 1, 2, \cdots, n$, denoting one type of disease, each record in T is associated with one or more disease labels in L. The problem of multi-label disease classification can be represented by a tuple of (m_i, S_i), where $S_i$is the class label for record m_i. S_i is a subset

Table 1
Mathematical equations for MLC.

Measure Name	Formula
Label Cardinality (LC)	$LC = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{L} y_j^{(i)}$ L = number of labels N = number of examples $y_j^{(i)}$ = is the set of labels of the ith instance
Label Density (LD)	$LD = \frac{LC}{L}$
Hamming Loss (HL)	$HL(h) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{Q}\lvert h(x_i \Delta \Upsilon_i)\rvert$ Υ_i = set of true labels of example x_i $h(x_i)$ = set of predicted labels for the same examples N = the total number of examples Q = the total number of possible class labels. Δ = symmetric difference between two sets.
Average Precision (AP)	$AP = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\Upsilon_i}\sum_{\lambda\in\Upsilon_i}\frac{\lvert\mathscr{L}_i\rvert}{rank_f(x_i,\lambda)}$ $\mathscr{L}_i = \{\lambda' \lvert rank_f(x_i,\lambda') \le rank_f(x_i,\lambda), \lambda' \in \Upsilon_i\}$
One Error (OE)	$OE = \frac{1}{N}\sum_{i=1}^{N}\left[\operatorname{argmax}_{\lambda\in Y} f(x_i,\lambda) \notin \Upsilon_i\right]$
Ranking Loss (RL)	$RL = \frac{1}{N}\sum_{i=1}^{N}\frac{\lvert D_i\rvert}{\lvert\Upsilon_i\rvert\,\lvert\overline{\Upsilon_i}\rvert}$ $D_i = \{(\lambda_m, \lambda_n) \lvert f(x_i,\lambda_m) \le f(x_i,\lambda_n)(\lambda_m,\lambda_n) \in \Upsilon_i \times \overline{\Upsilon_i}\}$ $\overline{\Upsilon}$ = complementary set of Υ *in* $\mathscr{L}$
Accuracy (Acc)	$Acc = \frac{1}{N}\sum_{i=1}^{N}\frac{\lvert h(x_i)\cap\Upsilon_i\rvert}{\lvert h(x_i)\cup\Upsilon_i\rvert}$
Micro-precision	$micro_{precision} = \frac{\sum_{j=1}^{Q} tp_j}{\sum_{j=1}^{Q}(tp_j + fp_j)}$ where tp_j, fp_j *and* fn_j are the number of true positives, false positives and false negatives for the label λ_j considered as a binary class. Q is the total number of possible class labels.
Micro-recall	$micro_{recall} = \frac{\sum_{j=1}^{Q} tp_j}{\sum_{j=1}^{Q}(tp_j + fn_j)}$
Micro-F1	$micro_{F1} = \frac{2\times micro_{precision}\times micro_{recall}}{micro_{precision} + micro_{recall}}$
Macro-Precision	$macro_{precision} = \frac{1}{Q}\sum_{j=1}^{Q}\frac{tp_j}{tp_j + fp_j}$
Macro-Recall	$macro_{recall} = \frac{1}{Q}\sum_{j=1}^{Q}\frac{tp_j}{tp_j + fn_j}$
Macro-F1	$macro_{F1} = \frac{1}{Q}\sum_{j=1}^{Q}\frac{2\times macro_{precision}\times macro_{recall}}{macro_{precision} + macro_{recall}}$

of L, which denotes $S_i \subseteq L$. The aim of the study is to build a classification model to predict 5 psychotic disease labels S_m' for every new physical record m_i'.

2.2. Machine learning evaluation measures

The degree of multi-label is measured by Label Cardinality (LC) and Label Density (LD). LC averages number of labels of the input examples in the dataset. For the PDD dataset, LC is 1.83 which signifies that each example has an average of more than 1 label associated with it. LD accounts for the number of labels with the average number of labels of the input examples (LC divided by the number of labels). The Lower the LD value, the slighter the number of occurrences of the label in the dataset. Lower LD shows that there are fewer samples corresponding to each label and hence the learning method needs to learn the label within those limited samples. Thus, the LD of 0.366 indicates that the average percentage of occurrence of each label in dataset is 36.6%. The Label Cardinality (LC) is the average number of labels of the observations in the Psychotic dataset while Label Density (LD) is LC divided by the Label set (L).

The machine learning evaluation measures are viewed from example-based, label-based and ranking-based perspectives. The examples-based are based on the average differences of the actual and the predicted sets of labels from the overall examples of the evaluated dataset. The label-based are used for predicting the performance of each label separately and then averaging the performance over all labels [23]. The ranking-based evaluation measures are used for comparing the predicted ranking of the labels and the ground truth ranking.

Table 2
Summary of the psychotic patients dataset.

S/N	Attribute	Type	Values	Size
1	Gender	Nominal	Female	267
			Male	233
2	Age Group	Nominal	<30	170
			30–60	278
			>60	52
3	History in Family	Nominal	Yes	269
			No	231
4	Religion	Nominal	Christianity	222
			Islam	219
			Others	59
5	Occupation	Nominal	Artisan	144
			Civil Servant	73
			Force	21
			Retired	46
			Student	120
			Unemployed	96
6	Hereditary	Nominal	Yes	221
			No	279
7	Status	Nominal	Married	281
			Single	219
8	Loss of Parent(s)	Nominal	Yes	298
			No	202
9	Divorce	Nominal	Yes	60
			No	440
10	Head Injury	Nominal	Yes	94
			No	406
11	Spiritual Consult	Nominal	Yes	347
			No	153
12	Insomnia (Class)	Nominal	Negative (N)	297
			Positive (P)	203
13	Schizophrenia (Class)	Nominal	Negative (N)	75
			Positive (P)	425
14	Vascular Dementia (VD) (Class)	Nominal	Negative (N)	154
			Positive (P)	346
15	Attention-Deficit/ Hyperactivity Disorder (ADHD) (Class)	Nominal	Negative (N)	282
			Positive (P)	218
16	Bipolar (Class)	Nominal	Negative (N)	299
			Positive (P)	201

The mathematical equations/formulas for comparing the evaluation measures of the MLC model used in this study are as presented in Table 1.

This study utilized three example-based (HL, Acc, EM), three-label based (Macro-F1 (example), Macro-F1(label), Micro-F1) and four ranking-based measures (OE, ZOL, AP, RL) evaluation measures. A threshold calibration method was recommended for this study in order to be able to minimize the difference in label cardinality between the training and the predictions of the test dataset [17]. The performance of the predictive model is informed based on the selection of an appropriate value of the threshold [23]. The probabilistic values of the Hamming Loss (HL), One Error (OE), Zero one Loss (ZOL), Ranking Loss (RL) ranges from 0 to 1. Hence, the performance of the model is superior to another if the value is 0 or the smallest value. For Accuracy (Acc), Average Precision (AP), Micro-F1 and Macro-F1, their values also range from 0 to 1 where large value indicates superior performance.

The Hamming Loss measures the ability of the algorithm to discriminate the symbols associated to the type of psychotic illness each patient has. Ranking loss measures the ranking of the labels according to the dominant disease of the patient.

Table 3
Summary of demographic variables of the patients in which the five psychotic ailments are present.

Variable		Insomnia	Schizophrenia	Vascular Dementia	ADHD	Bipolar
faNoily Status	Yes	111	222	196	116	110
	No	92	203	151	102	91
Religion	Christianity	84	193	148	98	82
	Islam	93	177	158	91	93
	Others	26	55	41	29	26
Genetic	Yes	88	180	151	92	86
	No	115	245	196	126	115
Marital Status	Single	55	205	156	124	53
	Married	148	220	191	94	148
Loss of Parent	Yes	118	250	206	123	116
	No	85	175	141	95	85
Divorce	Yes	37	41	37	24	37
	No	166	384	310	194	184
Injury	Yes	39	84	64	44	39
	No	164	341	283	174	162
Age	<30	44	173	125	104	42
	30–60	76	204	155	102	76
	>60	83	48	67	12	83
Gender	Male	87	203	162	140	85
	Female	116	222	185	78	116
Occupation	Artisan	56	131	101	57	53
	Civil Servant	43	62	49	22	44
	Force	6	21	14	14	6
	Retired	38	7	26	4	38
	Student	29	113	80	68	29
	Unemployed	31	91	77	53	31
Spiritual Consultation	Yes	153	291	302	151	152
	No	50	134	45	67	49
Total		**203 (40.6%)**	**425 (85.0%)**	**347 (69.4%)**	**218 (43.6%)**	**201 (40.2%)**

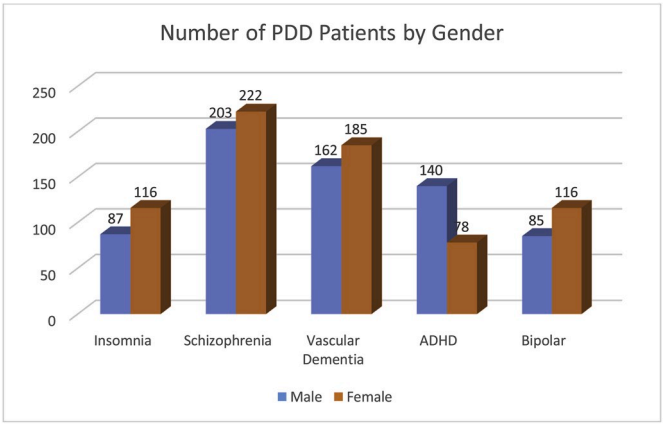

Fig. 2. Number of PDD patients by Gender.

3. Materials and methods

3.1. Data collection

The data were obtained from Yaba Psychiatry hospital, Yaba, Lagos state, Nigeria by Ref. [15]. It contained medical records of 500 psychotic patients, 16 variables (11 independent and 5 dependent variables). The information spans a period of five years (Jan. 2010–Dec. 2014). The summary of the dataset and the summary of the demographic variables of the patients with PDD are presented in Table 2 and Table 3 respectively.

3.2. Experiments

All the multi-label learning models and evaluation metrics were implemented with the experiment section of MEKA 1.9.2 [26], an open source multi-label machine learning suite based on WEKA [27]. The operating system is a 64-bit Windows 10 x64-based Processor an Intel Core i5-7200U CPU @2.50 GHz 2.70 GHz 8 GB. For all experiments performed, 10-fold cross-validation was applied to evaluate the performance systematically. All instances of the dataset were split into ratio of 90:10. The validation was iterated 30 times for the experiment, the averaged values of 300 runs are calculated for all models. All multi-label classifiers were trained with default parameters as advised by Ref. [28].

SVM, NB, LMT and NBTree base classifiers were used in the Problem Transformation (PT) and Ensemble methods in this study. For the PT method, we use BR, CC, LC, PCC, PS, FW and RT. And for the ensemble methods, we use RAkEL, RAkELd, EBR, ECC, EFW, EPCC, EPS, ELC and ERT. Having determined the best parameters values for each method on every dataset, the classifiers were trained using all available training examples and were evaluated by recognizing all test examples from the corresponding dataset.

4. Result and discussion

4.1. Description of the PDD dataset

The dataset distribution is almost normal with positive skewness of 0.495. Table 3 shows the demographic variables of the patients in which the five psychotic ailments are present. The patients consist of 40.6%, 85.0%, 69.4%, 43.6% and 40.2% suffering respectively from Insomnia, Schizophrenia, Vascular Dementia, ADHD and Bipolar. The female patients were more affected in all the psychotic ailments except in ADHD where the males are almost double (See Fig. 2). The Artisans were the

Table 4
Result of 90/10 percentage split.

Classifier	Example Based			Ranking Based				Label Based		
	Acc.	EM	HL	AP	OE	RL	ZOL	Macro-F1 (example)	Macro-F1 (label)	Micro-F1
BR -NB	0.4963	0.3360	0.2569	0.8802	0.4107	0.1351	0.6640	0.4362	0.6422	0.6448
BR -SVM	0.5513	0.3987	0.2332	0.7813	0.4073	0.2278	0.6013	0.4716	0.6287	0.6466
BR -LMT	0.5098	0.3153	0.2585	0.8855	0.4000	0.1316	0.6847	0.4872	0.6368	0.6480
BR-NBTree	0.4837	0.2933	0.2643	0.8579	0.4367	0.1562	0.7067	0.4482	0.6281	0.6343
CC - NB	0.4843	0.3273	0.2607	0.7026	0.4907	0.2901	0.6727	0.3932	0.6255	0.6232
CC - SVM	0.5513	0.3987	0.2332	0.7813	0.4073	0.2278	0.6013	0.4716	0.6287	0.6466
CC - LMT	0.5180	0.3520	0.2469	0.7569	0.4273	0.2458	0.6480	0.4489	0.6151	0.6280
CC - NBTree	0.4875	0.3253	0.2616	0.7219	0.4727	0.2748	0.6747	0.4138	0.6165	0.6182
FW –NB	0.5226	0.3460	0.2541	0.7704	0.4440	0.2214	0.6540	0.4630	0.6483	0.6551
FW -SVM	0.5450	0.3733	0.2440	0.8037	0.3953	0.2156	0.6267	0.4836	0.6315	0.6503
FW -LMT	0.5223	0.3440	0.2559	0.8016	0.4080	0.2133	0.6560	0.4817	0.6336	0.6468
FW -NBTree	0.5189	0.3353	0.2573	0.7940	0.4253	0.2003	0.6647	0.4742	0.6471	0.6546
LC –NB	0.5656	0.4060	0.2395	0.7739	0.4320	0.2201	0.5940	0.4985	0.6582	0.6710
LC -SVM	0.5434	0.3907	0.2493	0.7630	0.4367	0.2404	0.6093	0.4686	0.6330	0.6449
LC -LMT	0.5353	0.3893	0.2451	0.7577	0.4333	0.2534	0.6107	0.4441	0.6187	0.6312
LC -NBTree	0.5656	0.4060	0.2395	0.7739	0.4320	0.2201	0.5940	0.4985	0.6582	0.6710
PCC - NB	0.4884	0.3033	0.2577	0.8454	0.4487	0.1743	0.6967	0.4222	0.6424	0.6439
PCC - SVM	0.5513	0.3987	0.2332	0.7813	0.4073	0.2278	0.6013	0.4716	0.6287	0.6466
PCC - LMT	0.5080	0.2887	0.2536	0.8618	0.4100	0.1778	0.7113	0.4966	0.6386	0.6548
PCC - NBTree	0.4947	0.3027	0.2609	0.8416	0.4420	0.1844	0.6973	0.4478	0.6320	0.6393
PS – NB	0.5656	0.4060	0.2395	0.7739	0.4320	0.2201	0.5940	0.4985	0.6582	0.6710
PS - SVM	0.5434	0.3907	0.2493	0.7630	0.4367	0.2404	0.6093	0.4686	0.6330	0.6449
PS - LMT	0.5353	0.3893	0.2451	0.7577	0.4333	0.2534	0.6107	0.4441	0.6187	0.6312
PS - NBTree	0.5656	0.4060	0.2395	0.7739	0.4320	0.2201	0.5940	0.4985	0.6582	0.6710
RAkEL - NB	0.5395	0.3673	0.2476	0.8077	0.4040	0.1924	0.6327	0.4795	0.6568	0.6647
RAkEL - SVM	0.5419	0.3553	0.2479	0.8254	0.3980	0.1987	0.6447	0.4995	0.6482	0.6634
RAkEL - LMT	0.5368	0.3567	0.2507	0.8180	0.4027	0.2035	0.6433	0.4902	0.6442	0.6566
RAkEL - NBTree	0.5349	0.3607	0.2496	0.8165	0.4087	0.1874	0.6393	0.4847	0.6539	0.6638
RT – NB	0.3986	0.2080	0.3459	0.8807	0.4080	0.1362	0.7920	0.4704	0.5183	0.5227
RT - SVM	0.3939	0.0800	0.3363	0.8866	0.4040	0.1304	0.9200	0.4970	0.4972	0.5611
RT - LMT	0.4073	0.1987	0.3508	0.8826	0.4153	0.1333	0.8013	0.4824	0.4219	0.5306
RT - NBTree	0.3986	0.2080	0.3459	0.8807	0.4080	0.1362	0.7920	0.4704	0.5183	0.5227
EBR - NB	0.4951	0.3340	0.2563	0.8786	0.4133	0.1368	0.6660	0.4350	0.6425	0.6457
EBR - SVM	0.5355	0.3680	0.2488	0.8031	0.4020	0.2070	0.6320	0.4908	0.6372	0.6499
EBR - LMT	0.5374	0.3580	0.2491	0.8881	0.3960	0.1304	0.6420	0.4967	0.6422	0.6562
EBR - NBTree	0.5107	0.3273	0.2527	0.8774	0.4107	0.1380	0.6727	0.4639	0.6425	0.6509
ECC - NB	0.4989	0.3273	0.2615	0.7707	0.4553	0.2289	0.6727	0.4267	0.6418	0.6434
ECC - SVM	0.5456	0.3840	0.2481	0.8080	0.4033	0.2053	0.6160	0.4932	0.6386	0.6533
ECC - LMT	0.5366	0.3567	0.2557	0.8281	0.4100	0.1854	0.6433	0.5026	0.6458	0.6582
ECC - NBTree	0.5261	0.3540	0.2493	0.8495	0.4047	0.1627	0.6460	0.4713	0.6534	0.6616
EFW - NB	0.5169	0.3487	0.2521	0.8335	0.4040	0.1743	0.6513	0.4572	0.6478	0.6534
EFW - SVM	0.5405	0.3620	0.2519	0.8495	0.3953	0.1713	0.6380	0.4963	0.6420	0.6563
EFW - LMT	0.5402	0.3607	0.2508	0.8540	0.3947	0.1634	0.6393	0.4948	0.6427	0.6558
EFW - NBTree	0.5242	0.3460	0.2509	0.8632	0.4087	0.1551	0.6540	0.4751	0.6464	0.6555
ELC - NB	0.5513	0.3873	0.2453	0.8066	0.4287	0.1906	0.6127	0.4900	0.6534	0.6651
ELC - SVM	0.5544	0.3893	0.2483	0.8265	0.4067	0.1831	0.6107	0.4922	0.6537	0.6644
ELC - LMT	0.5482	0.3733	0.2497	0.8206	0.4133	0.1927	0.6267	0.4816	0.6499	0.6614
ELC - NBTree	0.5482	0.3820	0.2472	0.8150	0.4347	0.1888	0.6180	0.4874	0.6521	0.6628
EPCC - NB	0.4982	0.3267	0.2556	0.8703	0.4220	0.1448	0.6733	0.4274	0.6453	0.6466
EPCC - SVM	0.5456	0.3840	0.2481	0.8080	0.4033	0.2053	0.6160	0.4932	0.6386	0.6533
EPCC - LMT	0.5309	0.3427	0.2511	0.8888	0.3980	0.1297	0.6573	0.4997	0.6367	0.6534
EPCC - NBTree	0.5179	0.3340	0.2505	0.8744	0.4153	0.1418	0.6660	0.4677	0.6448	0.6537
EPS - NB	0.5513	0.3873	0.2453	0.8066	0.4287	0.1906	0.6127	0.4900	0.6534	0.6651
EPS - SVM	0.5544	0.3893	0.2483	0.8265	0.4067	0.1831	0.6107	0.4922	0.6537	0.6644
EPS - LMT	0.5482	0.3733	0.2497	0.8206	0.4133	0.1927	0.6267	0.4816	0.6499	0.6614
EPS - NBTree	0.5482	0.3820	0.2472	0.8150	0.4347	0.1888	0.6180	0.4874	0.6521	0.6628
ERT - NB	0.4000	0.2060	0.3424	0.8803	0.4067	0.1373	0.7940	0.4723	0.5206	0.5274
ERT - SVM	0.3919	0.1320	0.3408	0.8922	0.4000	0.1271	0.8680	0.4837	0.4277	0.5366
ERT - LMT	0.4058	0.2113	0.3500	0.8908	0.4027	0.1268	0.7887	0.4788	0.4033	0.5185
ERT - NBTree	0.4000	0.2060	0.3424	0.8803	0.4067	0.1373	0.7940	0.4723	0.5206	0.5274

major occupation group suffering from the ailments except in ADHD where the Students are more affected. Majority of the patients have had spiritual consultations in the course of their ailments. Those that are more than 60 years of age majorly suffer Insomnia and Bipolar. Schizophrenia and Vascular Dementia are more rampant in ages 30 to 60 group, while ADHD mostly affects less than 60 age groups. In terms of marital status, the married suffer more of the ailments, except in ADHD where the Students suffer more. A majority of the patients suffering from

Table 5
K–W test for Algorithm for each disorder.

	Algorithm	Mean Rank	Test Statistic	p-value	Decision
Insomnia	LMT	20.43	8.486	0.037	S
	NB	30.43			
	SVM	32.83			
	NBTREE	38.3			
Schizophrenia	NB	27.1	1.691	0.639	NS
	NBTREE	30.7			
	SVM	32.1			
	LMT	32.1			
VD	NB	23.27	6.798	0.079	NS
	LMT	29.53			
	NBTREE	30.9			
	SVM	38.3			
ADHD	NB	23.3	7.247	0.064	NS
	NBTREE	26.2			
	SVM	34.43			
	LMT	38.07			
Bipolar	LMT	21.83	5.508	0.138	NS
	NB	31.4			
	NBTREE	33.47			
	SVM	35.3			

all the ailments have lost their parents, are not divorced, without genetic disorder and injury. Most patients suffering from the ailments were of either Christianity or Islamic faith and with faNoily.

Multivariate (Pillar's Trace, Wilks' Lambda and Hoteling's Trace) test shows that Age, Occupation and Spiritual-Consultation were the only variables that are statistically significant across the five dependent variables (Insomnia, Schizophrenia, Vascular Dementia, ADHD and Bipolar). The univariate and the Post Hoc Tests show that Insomnia, Schizophrenia and ADHD are not significantly different statistically by ages "Less than 30" and "30–60" years, while they differ significantly for age 60. Bipolar differ significantly among the three age groups while there is no significant difference in Vascular Dementia by age group. On Occupation, Schizophrenia, Vascular Dementia, and ADHD differ significantly among the occupational groups. Vascular Dementia differs significantly only between the Retired and the Unemployed. Schizophrenia and ADHD differ significantly among the Retired people. Finally, Spiritual Consultation differs among the sufferers of Insomnia, Vascular Dementia and Bipolar.

4.2. *Machine learning experimental results*

The experimental results from the machine learning approach using 90/10 percentage split are presented in Table 4 for all the evaluation performance metrics, which are accuracy, exact match, Hamming loss,

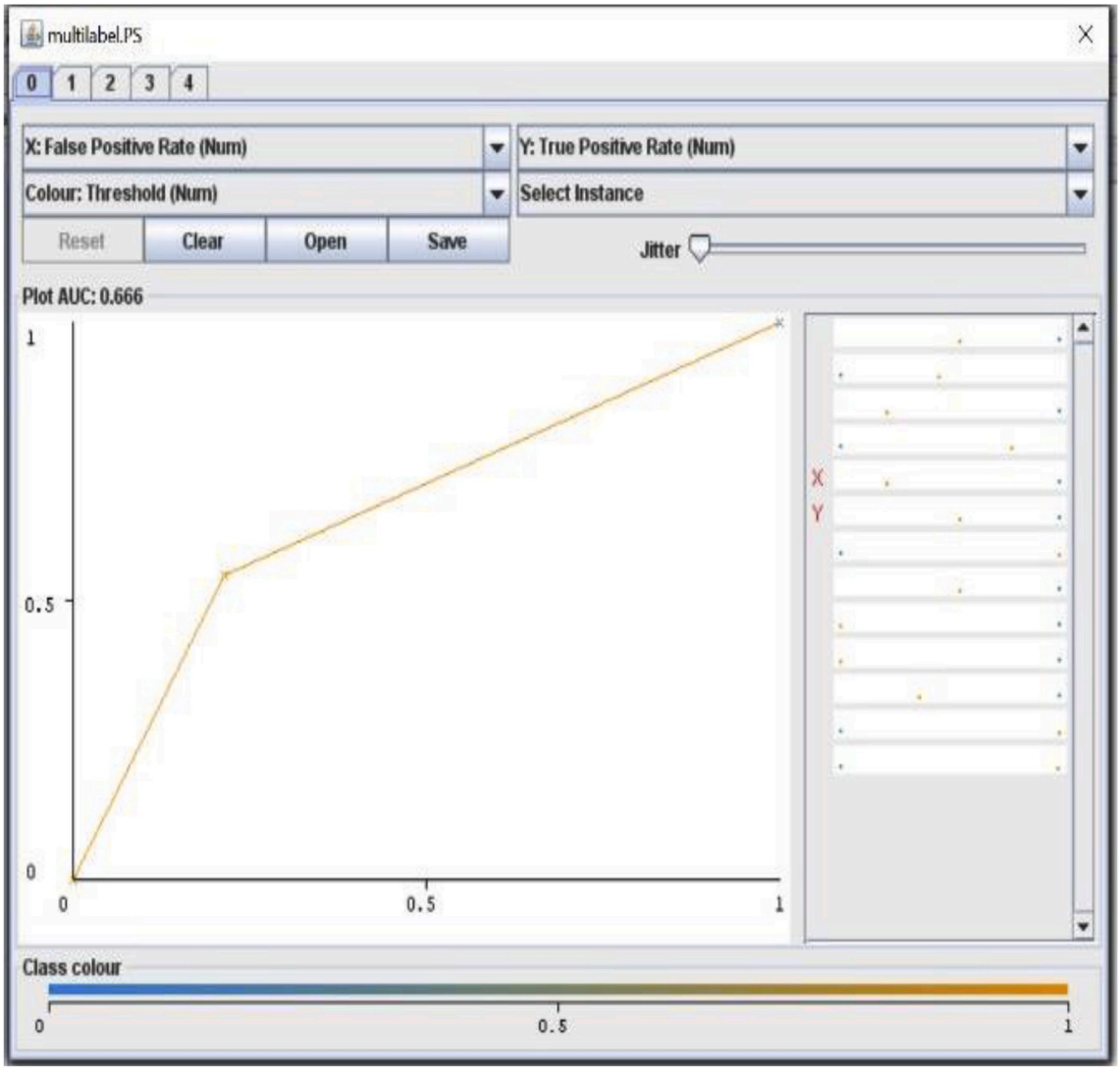

Fig. 3. ROC-AUC Graph for LC-NB on Insomnia.

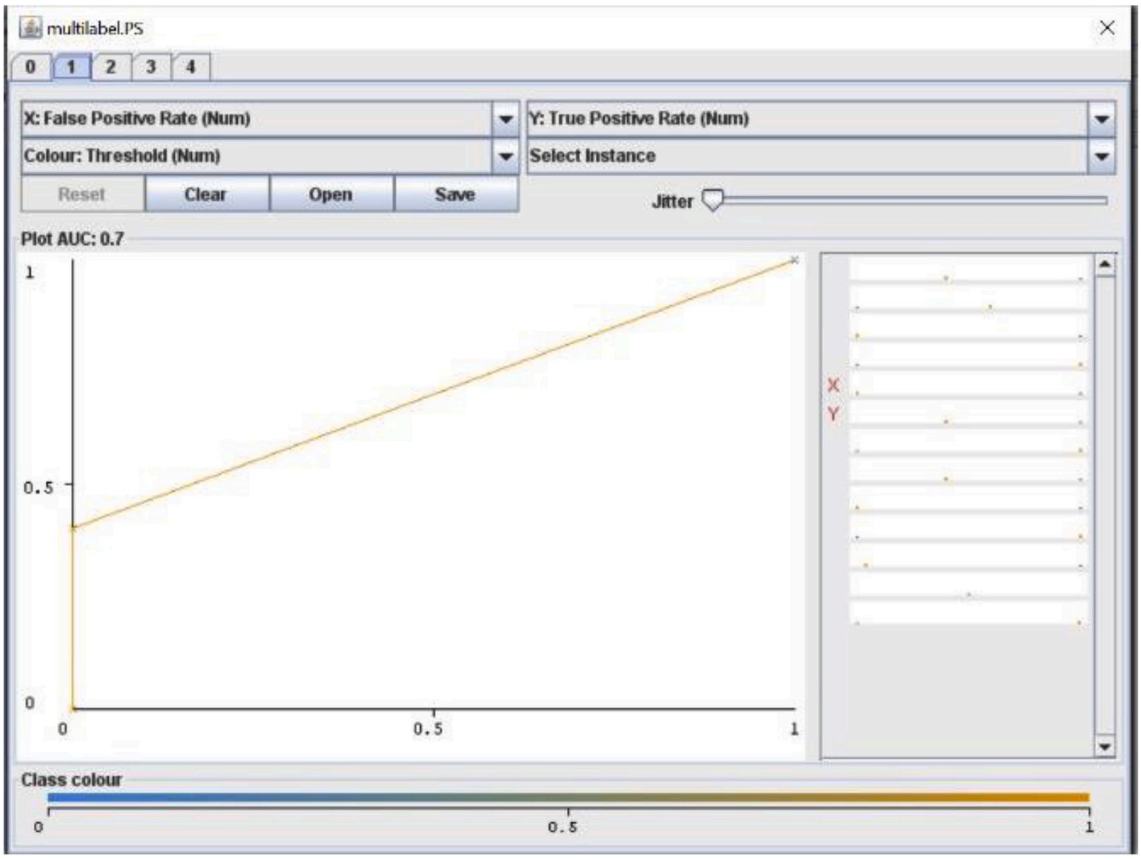

Fig. 4. ROC-AUC Graph for LC-NB on Schizophrenia.

average precision, one error, rank loss, zero one loss, F1(Macro by example), F1 (Macro by label) and Micro-F1. The result revealed that the following shows the best performance in terms of accuracy, exact match, Macro-F1(label) and Micro-F1 (LC and PS with NB and NBtree), Hamming Loss (BR- NBtree), Average precision (BR-LMT), one error (CC-NB), Rank Loss (CC-NB), Zero One Loss (BR-NBtree) and Macro-F1 (example) Rakel-SVM.

This study utilized different multi-label classification algorithms based on the fact that most researchers have reported in their findings that there is no single classification label software tool that is best [31]. The different multi-label classification algorithms classifiers employed were Logistic model tree (LMT), Naïve Bayes (NB), Naïve Bayes Tree (NBTree) and Support vector machines (SVM) for the problem transformation (PT) and the ensemble methods. For the PT method, BR, CC, LC, PCC, PS, FW and RT were used while RAkEL, RAkELd, EBR, ECC, EFW, EPCC, EPS, ELC and ERT were used for the ensemble methods. After determining the best parameter values for each method on every dataset, the classifiers were trained using all available training examples and were evaluated by recognizing all test examples from the corresponding dataset.

4.3. Statistical evaluation of the multi-label classification

The MLC were analyzed based on algorithm type, classifiers Measures type, and evaluation measures using the Kruskal-Wallis test (K–W).

The algorithms for One_Error and Macro_F1_Examples are statistically significantly different. All algorithms were not significantly different for RL, ZOL, ACC, AP, EM, HL, Macro_F1_label and Macro_F1. The algorithms are arranged based on the least mean rank (the best) for each evaluation measures (see Table 5). For instance, SVM is the best with One Error in the Ranking based measures, Accuracy with NB is the best in the Example-Based measures and Macro F1 example with NB is the best in the Label-based measures.

Of the classifiers, only Macro_F1_Example classifiers were not statistically significant from each other. The preferred classifier for One Error is EFW, ERT classifier for RL, LC for ZOL, RT for ACC, CC for AP, RT for EM, LC for HL, RT for Macro_F1_label and ERT for Macro_F1. The Ranked based measures are preferred to the Example-based and the Label-based measures.

On the evaluation measures, RL is the best. All evaluation measures are significantly different from each other. Table 5 shows that the algorithm differs statistically only for Insomnia (p-value < 0.05). LMT is the best algorithm for Insomnia and Bipolar while NB is the best for Schizophrenia, VD and ADHD.

As in Table 5, the classifiers performance in classifying each of the disorders was determined. The classifiers are statistically significantly different for Insomnia, VD and ADHD. The best classifier for Insomnia, VD, ADHD, and Bipolar is RT while Schizophrenia is best classified by CC.

4.4. Result based on MLC measures

Experimental results presented in Table 4 shows that for example-based (HL, Acc, EM), label-based (Macro-F1(example), Macro-F1 (label), Micro-F1) and ranking-based measures (OE, ZOL, AP, RL) evaluation measures, NB performs best as well as NBTree in most of the measures. And the best MLC are LC and PS. So, LC-NB shows the result of the MLC based on NB classifier. We summarize some observations briefly as follows:

i. LC performed as well as PS for all results. All of their obtained results were the same for all base classifiers. One of the reasons could be that LC treats each label combination as a single class in a multi-class learning scheme. The set of possible values of each

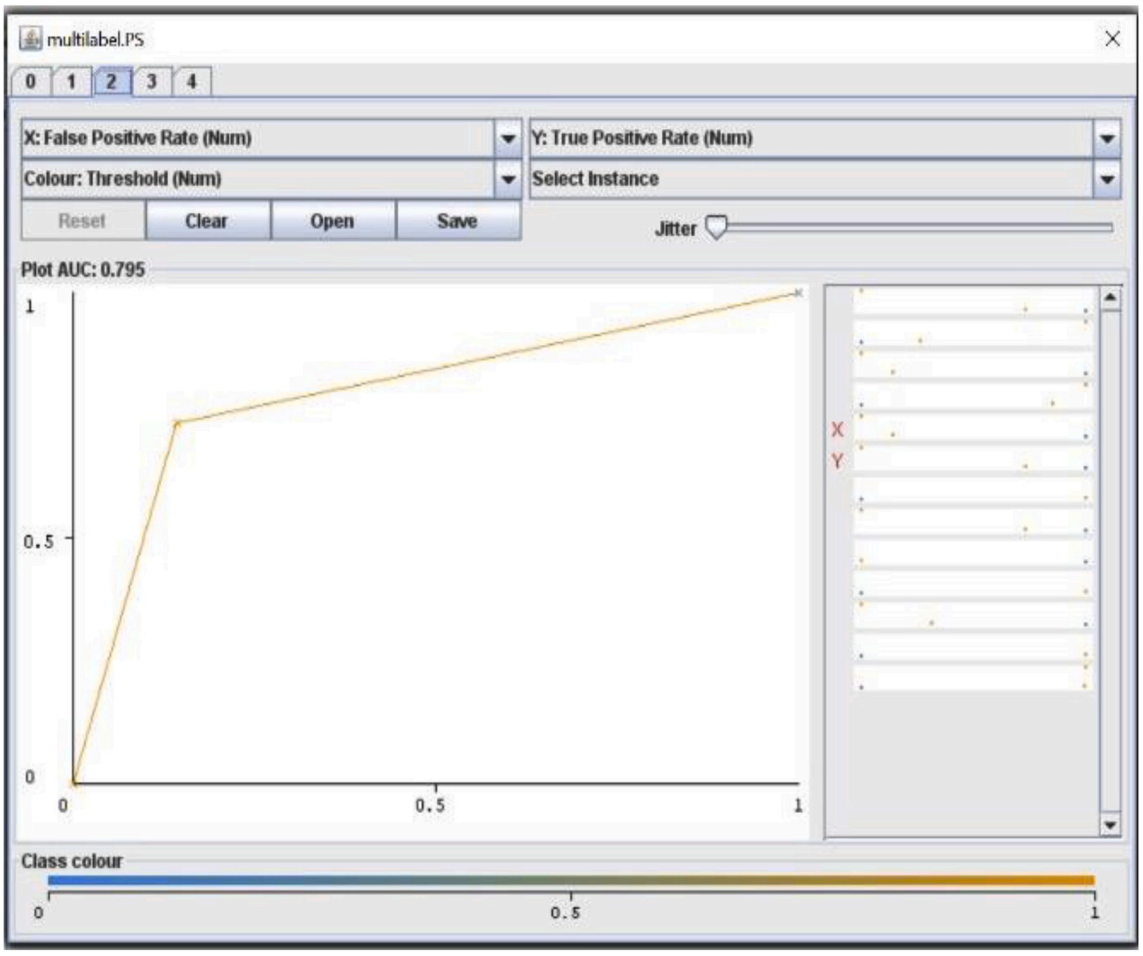

Fig. 5. ROC-AUC Graph for LC-NB on vascular Dementia.

class is the powerset of labels. Also, with PS, the idea is to reduce the number of unique class values that would otherwise need to be learned by LC.

ii. LC-NB, LC-NB-Tree, PS-NB and PS-NBTree significantly outperform other models for the example-based measure results. The averaged Acc. value (0.5656) and averaged EM value (0.4060) were the highest for all models.

iii. However, for HL, BR-SVM, CC-SVM and PCC-SVM performed best with the least value of 0.2332.

Investigation of the performances of some MLC models on 4 base classifiers in terms of the label-based measure of our PDD dataset shows that:

i The LC-NBTree and LC-NB obtained the same and highest values of 0.6582 and 0.6710 for Macro-F1 (label) and Micro-F1 respectively. For this measure, a higher value is better.

ii Macro-F1 (example)'s ECC-LMT, with a value of 0.5026, significantly outperforms all the other models. For this measure, a lower value is better.

Also, in Table 4, we investigate the performances of some MLC models on 4 base classifiers in terms of the ranking-based measure of our PDD dataset. A higher value is better for AP and ZOL measure while lower value shows a better model for OE and RL.

i. The experimental results (in Table 4) show that LC-NBTree and LC-NB obtained the same and highest value of 0.5940 for ZOL measure.

ii. For AP measure, ERT-SVM significantly outperforms all the other models with a value of 0.8922.

iii. For OE, EFW-LMT significantly outperforms all the other models with a least value of 0.3947, and

iv. For RL measure, ERT-LMT significantly outperforms all the other models with a least value of 0.1268.

4.5. Results based on ROC-AUC

This study also uses the area under the curve (AUC) and Receiver Operating Characteristics (ROC) curve to evaluate the model performance. ROC-AUC curve is a graphical representation for showing the trade-off between recall/sensitivity/true positive rate (TPR) and false positive rate (FPR) while precision-recall curve (PRC) is a graphical representation for showing the trade-off between precision and recall. The precision-recall curve (PRC) is not commonly used as the ROC-AUC curve in classification problem with balanced dataset [29,30]. In this study, the classes were represented as 0 (false positive) which shows that a patient is without the psychotic diseases among the patient with psychotic diseases and 1 (true positive) which shows that a patient is with the psychotic diseases among patients with psychotic diseases [31].

4.5.1. ROC-AUC curve

The optimal results of the ROC-AUC curve were selected in each of the different categories of the performance evaluation metrics on the multi-label classification model accuracy on PDD and presented in Figs. 3–7.

It is noted that the model for Vascular Dementia yields better performance with an area under the curve (AUC) of 79.5%. This indicates

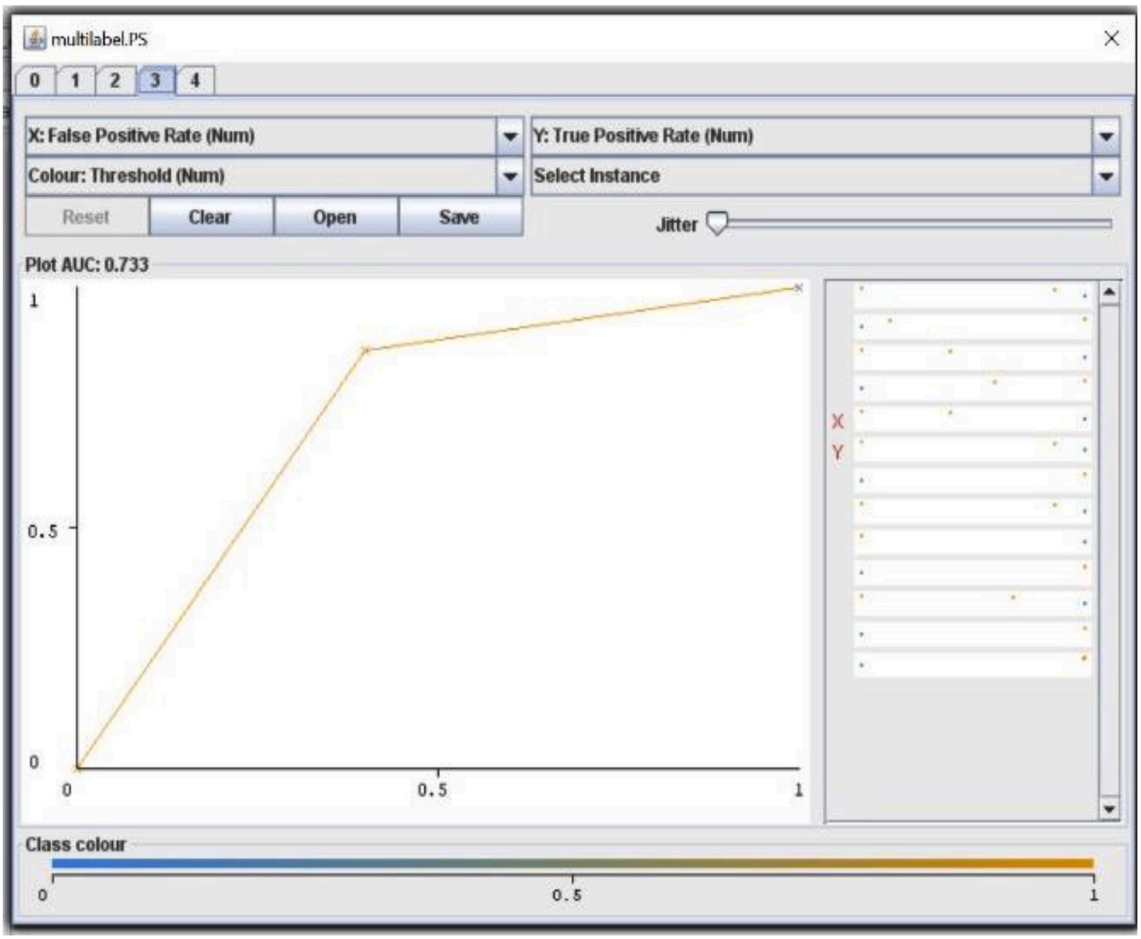

Fig. 6. ROC-AUC Graph for LC-NB on ADHD

that the model has more true positives and fewer false positives. On the other hand, the worst performance is exhibited by the model for Bipolar and Insomnia with 66.6% AUC.

4.5.2. Time-consumptions

The optimal results on the time-consumptions are based on the test time and total time. The total time of Naïve Bayes (0.02) is less as compared to SVM (1.107) and LMT (1.837) which means the Naïve Bayes outperforms the SVM and LMT while the test time of Naïve Bayes (0.004) and LMT (0.004) is less as compared to SVM (0.016) which means the Naïve Bayes and LMT outperform the SVM.

5. Conclusion

There is an urgent need for an early detection diagnostic decision support system for psychotic diseases to be integrated into the digital health technology that will guide against mental disorder in patients [32,33]. This study models the relationship between different psychotic diseases by means of the multi-label model in machine learning.

The MLC reported in this study is more efficient than the previous study carried out by Ref. [34] in terms of the ability to use Binary Relevance (BR) approach to solve m labels problem which cannot be handled by the use of only support vector machine on multi-label dataset; hence the inclusion of the statistical evaluation measures in this study.

Also, the experiments evaluated on multiple benchmark datasets were compared with Rank-SVM, ML-kNN, IBLR and Binary SVM. Our findings show that the proposed model significantly outperforms the other models at 5% significance level.

From medical perspective, the best choice to be considered on the recommended accuracy is the approach with fewer false positives. In this study Vascular Dementia has 20.5% false positives.

The study found that Schizophrenia has the highest classification accuracy while Bipolar has the lowest in the data split of 90-10.

The results obtained show that the Label Powerset (LC) and Pruned Sets (PS), MLC methods with Naïve Bayes (NB) and Naïve Bayes Tree (NBTree) consistently perform best in terms of the evaluation measures on the PDD dataset. Hence, either LC or PS methods are recommended for this dataset for model building. The classifiers are statistically significantly different for Insomnia, VD and ADHD. The best classifier for Insomnia, VD, ADHD, and Bipolar is RT while Schizophrenia is best classified by CC.

This study was challenged with its inability to generate a confusion matrix which can help to determine if there is an overlap in the true negative and false negative for the psychosis diseases.

Future study may explore the use of deep learning techniques on the multi-label classification to handle the confusion matrix that may enable us to determine if there will be an overlap between the false negative and true negative.

The use of machine learning and soft computing techniques will improve medical efficiency and effectiveness in mental healthcare systems for classification and prediction of psychotic diseases especially in developing countries.

Declaration of competing interest

The authors declared that there is no conflict of interest.

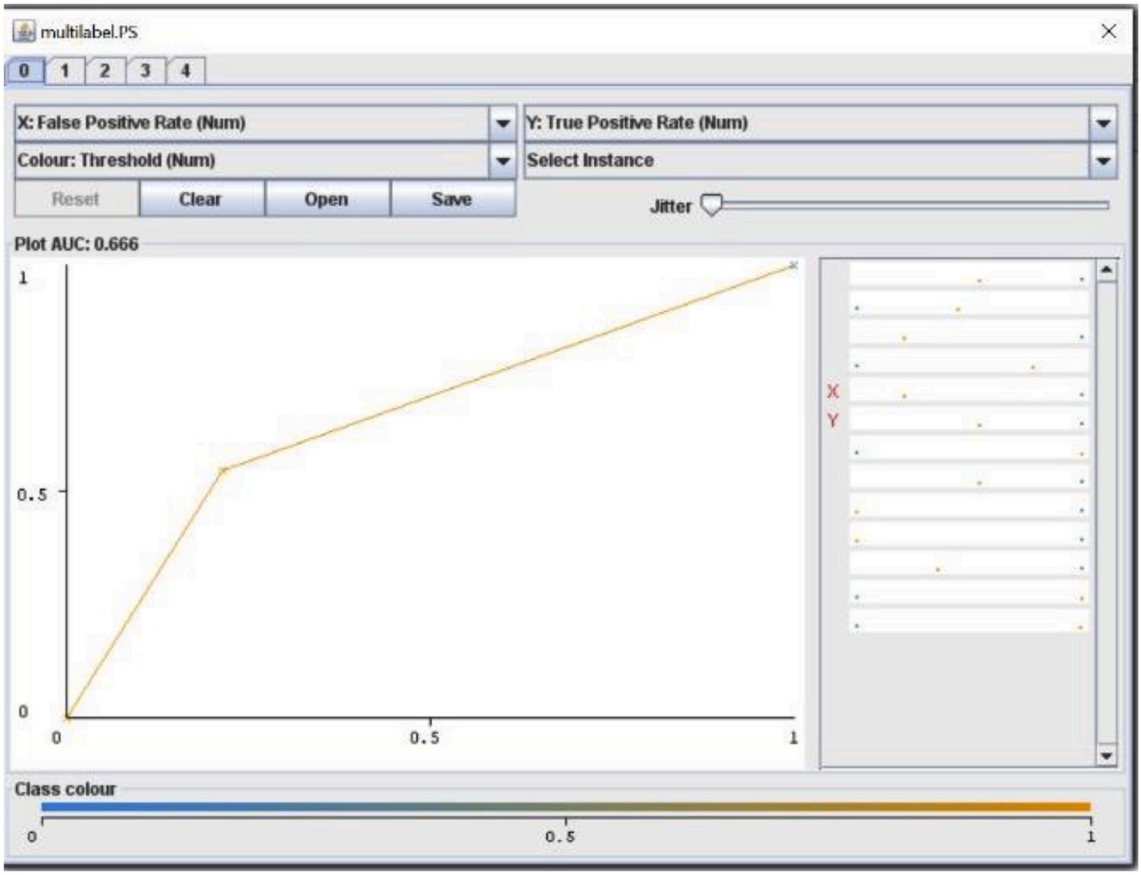

Fig. 7. ROC-AUC Graph for LC-NB on Bipolar.

Acknowledgement

We would like to appreciate the Yaba Psychiatry hospital, Yaba, Lagos state, Nigeria team and Adejumo, A. O., Ikoba, N. A., Suleiman, E. A., Okagbue, H. I., Oguntunde, P. E., Odetunmibi, O. A., Job, O for making the dataset available for research purpose. We will like to also appreciate the anonymous reviewers for their constructive criticism.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.imu.2020.100326.

References

[1] World Health Organization (WHO). The ICD-10 classification of mental and behavioural disorders. Geneva. 1993.
[2] Markov IS, Berrios GE. The meaning of insight in clinical psychiatry. Br J Psychiatr 1992;160:850–60.
[3] Jacobsen SR. Vascular dementia: risk factors, diagnosis and Treatment. Neuroscience research progress series. United Kingdom: Nova Science Publishers Inc.; 2011. March.
[4] Anastopoulos AD, Guevremont DC, Shelton TL, DuPaul GJ. Parenting stress among families of children with attention deficit hyperactivity disorder. J Abnorm Child Psychol 1992;20(5):503–20.
[5] Tsoumakas G. I. Katakis multilabel classification: an overview. Int J Data Warehous Min 2007;3(3):1–13.
[6] Ojeme B, Mbogho A. Selecting learning algorithms for simultaneous identification of depression and comorbid disorders. In: 20th international conference on knowledge based and intelligent information and engineering systems, vol. 96. Elsiever; 2016. p. 1294–303. Retrieved from, www.sciencedirect.com.
[7] Park SH, Furnkranz J. Efficient pairwise classification. In: 18th European conference on machine learning; 2007. p. 658–65.
[8] Dembczynsky K, Cheng W, Hullerme E. Bayes optimal multi-label classification via probabalistic classifier chains. In: 27th international conference on machine learning (ICML'10); 2010.
[9] Read J, Blieza C, Larranaga P. Multi-dimensional classification with super-classes. IEEE Trans Knowl Data Eng 2014;26(7):1720–33. https://doi.org/10.1109/tkde.2013.167.
[10] Breiman L. Bagging predictors. In: Quinlan R, editor. Machine learning, vol. 24; 1996. p. 123–40. Retrieved February 17, 2018, from, https://link.springer.com/content/pdf/10.1023/A:1018054314350.pdf.
[11] Read J, Pfahringer B, Holmes G, Frank E. Classifier chain for multi-label classification. Mach Learn 2011;85(3):333–59.
[12] Read J, Reutemann P, Pfahringer B, Holmes G. MEKA: A Multi-label / Multi-target Extension to WekaKegl B, editor. Journal of Machine Learning Research 2016;17 (21):1–5. Retrieved from, http://jmlr.org/papers/v17/12-164.html.
[13] Li R, Zhao H, Lin Y, Maxwell A, Zhang C. Multi-label classification for intelligent health risk prediction. IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2016:986–93.
[14] Cheng W, Hüllermeier E. Combining instance-based learning and logistic regression for multilabel classification. Mach Learn 2009;76:211–25. https://doi.org/10.1007/s10994-009-5127-5.
[15] Yang Y, Gopal S. Multilabel classification with meta-level features in a learning-to-rank framework. Mach Learn 2012;88:47–68. https://doi.org/10.1007/s10994-011-5270-7.
[16] Read J, Pfahringer B, Holmes G, Frank E. Classifier chains for multi-label classification. In: al WB, editor. 20th European conference on machine learning PKDD, LNAI 5782, Part II; 2009. p. 254–69.
[17] Zaragoza J, Sucar L, Bielza C, Larranaga P. Larranaga Bayesian chain classifiers for multidimensional classification. In: Twenty-second international joint conference on artificial intelligence (IJCAI); 2011. p. 2192–7.
[18] Zhang ML, Zhou ZH. ML-kNN: a lazy learning approach to multi-label learning. Pattern Recogn 2007;40(7):2038–48.
[19] Schapire RE, Singer Y. Boostexter: a Boosting - based system for text categorization. Mach Learn 2000;39(2/3):135–68.
[20] Clare A, King RD. Knowledge discovery in multi-label phenotype data. In: 5th European conferenceon PKDD; 2001. p. 42–53.
[21] Blockeel J, Raedt LD, Ramon J. Top - down induction of clustering trees. In: 5th international conference on machine learning; 1998. p. 55–63.
[22] Madjarov G, Kocev D, Gjorgjevikj D, Dzeroski S. An extensive experimental comparison of methods for multi-label learning. Pattern Recogn 2012;45: 3084–104. https://doi.org/10.1016/j.patcog.2012.03.004.
[23] Tsoumakas G, Vlahavas I. Random k-labelsets: an ensemble method for multilabel classification. In: 18th European conference on machine learning; 2007. p. 406–17.
[24] Kocev D. Ensembles for predicting structured outputs. IPS Jožef. Stefan, Ljubljana, Slovenia. 2011.

[25] Read J, Pfahringer B, Holmes G. Multi-label classification using ensembles of pruned sets. In: International conference on data mining (ICDM'08). Pisa, Italy; 2008.
[26] Adejumo AO, Ikoba NA, Suleiman EA, Okagbue HI, Oguntunde PE, Odetunmibi OA, Job O. Quantitativ eexploration of factors influencing psychotic disorder ailments in Nigeria. Data in Brief 2017;14:175–85. https://doi.org/10.1016/j.dib.2017.07.046.
[27] Witten IH, Frank E, Hall MA. Data mining practical machine learning tools and techniques. third ed. Morgan Kaufmann; 2011. Retrieved 2018.
[28] Tahir MA, Kittler J, Mikolajczyk K, Yan F. Improving multilabel classification performance by using ensemble of multi-label, classifiers. In: El Gayar N, Kittler J, Roli F, editors. MCS. Springer-Verlag Berlin Heidelberg; 2010. p. 11–21.
[29] Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PloS One 2015;10(3):e0118432.
[30] Xiong X, Li Q, Yang WS, Wei X, Hu X, Wang XC, Zhu D, Li R, Cao D, D, Xie P. Comparison of swirl sign and black hole sign in predicting early hematoma growth in patients with spontaneous intracerebral hemorrhage'. Med Sci Mon Int Med J Exp Clin Res: Int. Med.J.Exp. Clin. Res. 2018;24:567.
[31] Akinnuwesi BA, Adegbite AB, Adelowo F, Ima-Edomwonyi U, Fashoto SG, Amumeji OT. Decision support for diagnosing Rheumatic-Musculoskeletal Disease using Fuzzy cognitive technique. In: Informatics in medicine unlocked, vol. 18. Elsevier; 2020. p. 100279.
[32] Uzoka FME, Akinnuwesi BA, Amoo T, Debele F, Fashoto SG, Nwafor-Okoli C. An expert system for malaria diagnosis using the fuzzy cognitive map engine. In: Cunningham P, Cunningham M, editors. International conference of IEEE-IST africa. International Information Management Corporation (IIMC); 2018. p. 1–13. Retrieved from, www.IST-Africa.org/Conference2018.
[33] Batra S, Baker R, Wang T, Forma F, DiBiasi F, Peters-Strickland T. Digital health technology for use in patients with serious mental illness: a systematic review of the Literature. Medical Devices (Auckl) 2017;10:237–51.
[34] Tran T, Kavuluru R. Predicting mental conditions based on "history of present illness" in psychiatric notes, with deep neural networks. J. Biomed. Inf.Vol. 2017;75(Supplement):S138–48. November 2017, Elsevier.